5

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: **84302723.6**

(22) Date of filing: **24.04.84**

(51) Int. Cl.³: **C 12 N 15/00**
C 12 P 21/02, C 12 N 1/18
//(C12N1/18, C12R1/865)

(30) Priority: **25.04.83 US 488323**
06.06.83 US 501352
20.06.83 US 506098

(43) Date of publication of application:
31.10.84 Bulletin 84/44

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: **GENENTECH, INC.**
460 Point San Bruno Boulevard
South San Francisco California 94080(US)

(72) Inventor: **Singh, Arjun**
735 Cape Breton Drive
Pacifica California 94044(US)

(74) Representative: **Armitage, Ian Michael et al,**
MEWBURN ELLIS & CO. 2/3 Cursitor Street
London EC4A 1BQ(GB)

(54) **Process for expressing heterologous protein in yeast, expression vehicles and yeast organisms therefor.**

(57) The isolation of the yeast α-factor genes is described. The promoter and signal peptide portions are isolated and joined to DNA coding for proteins heterologous to yeast in a plasmid which is used to transform yeast cells. The yeast expresses the heterologous DNA and processes and secretes the heterologous protein.

EP 0 123 544 A2

0123544

5

10 PROCESS FOR EXPRESSING HETEROLOGOUS PROTEIN IN
    YEAST, EXPRESSION VEHICLES AND YEAST ORGANISMS
                      THEREFOR

15

20

        This application is related to US applications Serial No.
    06/438,236 (EP 88632), and Serial No. 06/488,337 filed 25 April 1983
25 (European application filed on even date), the disclosures of
    which are hereby incorporated by reference.

30
    Field of the Invention

        This invention is directed generally to recombinant DNA
    technology utilizing yeast host systems and expression vehicles that
35 produce, process and secrete heterologous protein as discrete

    0384L

product unaccompanied by interfering amounts of unwanted presequence or other artifact of expression.

Proteins that are secreted through the cell membrane of the parent cell are ordinarily produced in the cell as a "pre"-protein. In that form, the protein is fused to an additional polypeptide sequence which presumably assists its secretion and localization. This additional protein, referred to as a "signal" polypeptide, is believed to be clipped from the secreted "mature" protein during the secretion process. Although the signal peptides of pre-proteins share some similarities, their primary structures differ considerably. The signal peptides even for a given organism exhibit this variation. For example, the signal for human growth hormone is substantially different from the signal for human insulin. This suggests that each protein has evolved with a signal sequence which is particularly well suited for translocation of that particular protein through a cell membrane.

This invention is based upon the discovery that a substantially mature protein is produced and often secreted by yeast when the DNA coding for the heterologous protein is operably attached to the DNA sequence of the promoter and/or signal peptide encoding portions of the yeast α-factor gene. (It will be apparent from the disclosure herein that yeast harbors at least two alpha factor genes. The use of "the alpha factor gene" is intended to include all such functional genes.) Thus, in a primary aspect, this invention is directed to the means and methods of obtaining useful quantities of heterologous protein from the medium of a yeast culture containing viable cells harboring expression vehicles containing DNA encoding the desired protein, wherein the DNA coding for this heterologous protein is operably connected to a DNA sequence comprising the promoter and/or signal portion of the yeast α-factor gene. Of enormous advantage is the enablement, by this invention, of obtaining useful, discrete protein product in the cell culture medium by expression of heterologous DNA in an easily modified plasmid.

The alpha factor of yeast contains a "pre-pro" sequence which is ordinarily removed from the α-factor upon the completed act of secretion. Operationally, therefore, the pre-pro sequence functions as a signal sequence in the process of secretion into the medium as

5 will be further explained below. It is clear from the results obtained herein that the peptide which comprises the thus-defined signal sequence of alpha factor fused to a heterologous protein is successfully processed by the yeast organism so as to result in.the secretion of the mature heterlogous protein into the surrounding

10 medium. Therefore, the advantages obtained by use of this "pre-pro" signal are realized whether or not the expression of the signal/heterologous protein gene sequence is under the control of the alpha factor promoter or under the control of other promoters which are functional in yeast. Similarly, the results obtained

15 demonstrate that the alpha factor promoter is effective in expressing the heterologous gene, and that such expression could be obtained without the intermediate insertion of the signal sequence into the expression vehicle. Accordingly, this invention is directed to the use of alpha factor promoter qua promoter in yeast

20 systems for the expression of heterologous peptides and to the use of the alpha factor signal qua signal as a means for effecting processing and secretion of heterologous proteins produced as a result of expression in yeast.

25 The publications and other materials referred to herein to illuminate the background of the invention, and in particular cases, to provide additional detail respecting its practice, are incorporated herein by reference, and for convenience, are numerically referenced and grouped in the appended bibliography.

30

Background of the Invention

Yeast organisms naturally transport a small number of certain
35 homologous proteins to, and sometimes through, the plasma membrane

0384L

as an essential contribution to cell surface growth and cell
metabolism. As the cell buds as an incident of reproduction
preparatory to formation of a daughter cell, additional proteins are
required for formation of cell wall and plasma membrane as well as
5   for metabolism. Some of these proteins must find their way to the
site of function; hence, a secretory pathway is believed to exist
(1). Certain homologous proteins involved in the above processes
are formed by translation by ribosomes attached to the endoplasmic
reticulum. Homologous proteins are those normally produced by the
10   yeast species and required for its viability. Once formed, they
migrate by transfer to Golgi apparatus, thence within vesicles to
plasma membranes where some associate, or to some extent, penetrate
into the space between the plasma membrane and the cell wall. A
small number of homologous proteins seems to be exported completely
15   through the cell wall, such as α-factor and killer toxin (2,3).

Again, the bud region of the cell seems to be the site of
attraction for the vesicles and by their fusion to the inner surface
of the bud they contribute to the overall growth of the plasma
20   membrane, and presumably, the cell wall (4,5,6). It is
controversial still whether glycosylation of the protein may assist,
or is implicated, in the so-called secretory process. Further, by
definition "secreted" proteins are believed to have a signal
prepeptide, postulated to be associated with the transport or
25   incorporation process at the membrane surface. However, the precise
mechanism involved in the overall secretory process is not fully
understood.

It was contemplated that recombinant DNA technology could
30   provide valuable assistance in answering the open questions about
the secretory process in yeast organisms and, given its proven
applicability in enabling such, and other, organisms to produce
copious quantities of heterologous polypeptide products endogenously
(See, e.g., 7 to 17), in achieving appropriate manipulation of the
35   yeast host so as to direct the secretion of heterologous protein in

0384L

discrete, mature form. This has, in fact, been achieved and is the subject of US 438,236 (EP 88632),

supra. In that application is described the discovery that a heterologous protein, initially expressed as a pre-protein with its native signal or hybrid thereof, can be processed and secreted by yeast as a mature protein.

Summary of the Invention

This invention is based on the discovery that yeast organisms can be caused to produce, process and secrete protein that is normally heterologous to the yeast organism and not required for its viability, such that the protein can be obtained from the medium supporting the viable and reproducing yeast cells and in discrete form substantially unaccompanied by unwanted peptide presequence or other artifact of expression. For this purpose, a DNA sequence encoding the desired, heterologous protein is linked to the DNA sequence encoding the non-native (to the protein) signal sequence of yeast α-factor. Suitable yeast cells are transformed with expression vehicles harboring such DNA encoding a heterologous protein operably connected to the such DNA coding for the α-factor signal (pre-pro) peptide and a promoter. Upon expression of the sequence encoding the heterologous protein together with that encoding α-factor signal peptide, the expression product is processed and the mature heterologous protein is exported into the medium of the cell culture, from which it can be removed with relative ease, without need to disrupt the viable yeast cells. It is thus recovered in otherwise substantially mature form for use, without the need to remove unwanted presequence or certain other artifacts of expression (e.g., the methionine attached to the otherwise first N-terminus amino acid which is an expressional consequence of the AUG translational start signal codon). Thus, the medium can be obtained in a form substantially free of viable or disrupted (i.e., lysed or otherwise broken) cells and, since it contains the desired product, is susceptible to more easily employed

0384L

purification techniques. Such product, after purification, is fit
for use as intended. For example, human leukocyte interferon
product finds use as a human antiviral and/or antitumor agent (See,
generally, 7 to 17).

In summary, the present invention comprises the use of yeast
alpha factor signal sequences and/or promoter to produce a protein
normally heterologous to a yeast organism and not required for its
viability, in discrete form unaccompanied by any substantial peptide
presequence or other artifact of expression, as a product of yeast
expression, processing and secretion. Further, this invention
provides yeast cultures capable of producing such protein and
resultant yeast culture media containing such protein as product.
More specifically, the invention is directed to a process for
producing heterologous proteins in yeast, and the expression
vehicles and organisms employed in this process, wherein the alpha
factor promoter is used to effect the expression of the foreign
gene. Further, the invention is directed to the use of the signal
(pre-pro) sequence for alpha factor to effect the processing and
secretion of an expressed foreign protein, to a recombinant
expression vehicle effectively harboring the alpha factcor DNA
sequences and to the cells transformed with such vehicles.

By the term "heterologous protein" as used herein is meant
protein that is not normally produced by or required for viability
of a yeast organism. This term contemplates the functional
insertion of DNA encoding such protein, via recombinant DNA
technology, into an expression vehicle, in turn used to transform a
yeast organism host. Functional insertion of DNA denotes the
insertion of DNA encoding the heterologous protein into an
expression vector under control of the α-factor promoter and/or
connected to the DNA sequence coding for the α-factor signal to
obtain a hybrid preprotein, i.e., one which comprises the α-factor
signal peptide fused to the heterologous protein. Examples of such
heterologous protein are hormones, e.g., human growth hormone,

0384L

bovine growth hormone, etc.; lymphokines; enzymes; interferons,
e.g., human fibroblast, human immune and human and hybrid leukocyte
interferons, bovine interferons etc.; viral antigens or immunogens,
e.g., foot and mouth disease antigens, influenza antigenic protein,
5   hepatitis core and surface antigens, etc.; factors incidental to
growth, e.g. human insulin-like growth factor (IGF-1 and IGF-2),
epidermal growth factor (EGF) and nerve growth factor (NGF) and
various other polypeptides, e.g., rennin, human serum albumin, human
insulin, various glycoproteins, etc.

10

"Secretion" as used herein means exportation of product through
the plasma membrane and at least into or through the cell wall of
the yeast organism into the medium supporting the cell culture. In
this connection, it will be understood that in some instances,
15   "secreted" product associates in some manner with the cell wall,
perhaps necessitating a different purification procedure or a
modification of the structure and function of the yeast host.

"Processing" means the cellular cleavage of the α-factor signal
20   peptide from the mature protein so as to produce the heterologous
protein unaccompanied by any substantial portion of the signal
sequence or by extraneous peptide in--so-called discrete--mature
form. By "extraneous" peptide is included peptide artifacts of
expression such as methionine. Processing admits of cleavage of the
25   signal polypeptide at a locus inconsequentially removed from the
precise point of signal peptide union with mature protein.

Brief Description of the Drawings

30

Fig. 1 illustrates the structure of pools of synthetic
oligonucleotides used as hybridization probes to isolate the gene
for α-factor.

35   Fig. 2 illustrates the results of electrophoresis of DNA

0384L

fragments obtained using the probes of Fig. 1.

Figs. 3 and 4 are the nucleotide sequences of α-factor genes.

5   Fig. 5 illustrates the scheme for joining the gene for human interferon D with the gene for the α-factor promoter and signal sequence.

Fig. 6 illustrates the scheme for construction of a yeast
10 expression plasmid for expression of human interferon D (IFN-α_1).

Fig. 7 depicts the protein and DNA sequence at the junction of the α-factor signal sequence and the modified IFN-α_1 gene.

Fig. 8 shows the levels of IFN-α_1 in the medium and cell
15 extracts of a culture of a yeast transformant expressing IFN-α_1.

Fig. 9 illustrates the scheme for construction of a yeast/E. coli shuttle vector for expression of heterologous genes
20 using the α-factor promoter and signal polypeptide gene sequences.

Fig. 10 illustrates the assembly of a yeast/E. coli plasmid for expression of tissue plasminogen activator.

25   Fig. 11 depicts the construction used to effect the production of mature heterologous proteins (human interferon   illustrated) as a product of expression with partial alpha factor signal sequence, processing of the alpha factor componment and secretion of the mature protein into the supporting medium.

30   Fig. 12 illustrates the degree of consensus between the MFα1 and MFα2 polypeptides of figures 3 and 4.

35

0384L

Detailed Description of a Preferred Embodiment

The yeast <u>Saccharomyces cerevisiae</u> secretes only a limited
number of proteins into the culture medium. One of the proteins
that is found in the medium is α-pheromone or α-factor (2). Duntze
and coworkers (18, 19) first determined that the α-factor is a
family of four oligopeptides of 12-13 amino acid residues having the
basic sequence
H_2N-(Trp)-His-Trp-Leu-Gln-Leu-Lys-Pro-Gly-Gln-Pro-Met (or MetSO)-Tyr-COOH.
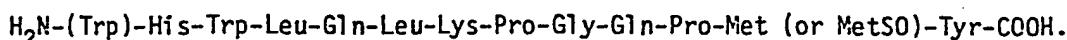
Figure 3 shows the location of the four peptides in the
unprocessed product of one of the α-factor genes, MFα-1. Only the
"boxed" segments are secreted into the medium, the remaining
sequences are not. It is not clear how much of the remaining
sequence is "true" signal (pre) sequence, which is at least
partially processed to effect secretion, and how much is "pro"
sequence in the sense of a traditional precursor protein (e.g.,
prorennin, proinsulin.)

Similarly, only the "boxed" portions in product of the <u>MFα2</u>
gene shown in Figure 4 are secreted, and the nature of the remaining
sequences can be described analogously to that of those in <u>MFα1</u>.

During the reduction of this invention to practice, another
group (44) succeeded in isolating and sequencing one of the genes
for α-factor (<u>MFα1</u>) by methods different from those disclosed
herein. As described below, two α-factor genes were isolated by us
and expression vectors in which the DNA sequence for the promoter
and signal peptide of α-factor was inserted in tandem with the DNA
sequence coding for heterologous protein were constructed from one
of them.

A. <u>Bacterial and Yeast Strains</u>

<u>E</u>. <u>coli</u> K-12 strain 294 (<u>endA</u> <u>thi</u>⁻hsr⁻hsm⁺)(ATTC

0384L

31446)(22) was used for bacterial transformations. Yeast strain
20B-12 (α, trp1 pep4) deposited without restriction in the American
Type Culture Collection, ATCC No. 20626, on March 5, 1982 was used
as yeast host.

B. Growth Media

The routine yeast growth medium contained 1 percent
Bacto-yeast extract, 2 percent Bacto-peptone and 2 percent
dextrose. Yeast minimal medium contained 0.67 percent Bacto-yeast
nitrogen base without amino acids, 2 percent dextrose and 3 percent
gar. The minimal medium supplemented with 1M sorbitol was used for
yeast transformations. Bacterial growth medium was LB (25) which
was supplemented with 20 µg/ml ampicillin when used for
transformation. S-agar plates used for colony screening contained
per liter: 32g tryptone, 5g NaCl, 15g Difco agar and 0.2g NaOH to
which ampicillin or chloramphenicol was added as indicated.

C. Transformations

E. coli 294 was transformed using a published procedure
(23). Yeast were transformed essentially as described (21, 24).

D. Enzymes and DNA Preparations

Restriction enzymes were purchased from New England Biolabs
and Bethesda Research Laboratories and were used according to
manufacturer's recommendations. T4 DNA ligase was from New England
Biolabs and was used in 20mM Tris-HCl (pH 7.5), 10mM $MgCl_2$, 10mM
dithiothreitol, 1mM ATP at 14°C. Calf alkaline phosphatase was
purchased from Boehringer Mannheim and was used in 100mM NaCl, 50mM
Tris-HCl (pH 7.4), 10mM $MgSO_4$, 1mM 2-mercapto-ethanol at 37°C.

0384L

Plasmid DNAs were prepared by the cleared lysate method (29) and were purified by Bio-Rad Agarose A-50 column chromatography. Small amounts of plasmid DNAs from individual E. coli transformants were prepared by a quick-screening procedure (20). DNA restriction fragments were isolated by electroelution from a 1 percent agarose gel followed by phenol/chloroform extraction and ethanol precipitation. Oligo-deoxynucleotide probes were prepared by the phosphotriester method (41).

E. Design of the Hybridization Probe

The 15-mer oligonucleotide probes for the α-factor gene were designed on the basis of the amino acid sequence of the pheromone (19) and yeast codon usage frequencies. The rationale is outlined in Fig. 1 where the last 5 amino acids of the α-factor and all the possible codons and their usage frequencies are given. (The codon usage is the total of 2 different glyceraldehyde-3-phosphate dehydrogenase clones (30, 31) and of alcohol dehydrogenase I.) The codon usage for these and other genes has recently been summarized (45). As can be seen from Fig. 1, virtually all possible sequences coding for the 5 amino acids are included in the oligonucleotide sequence 5'-GG$_C^T$CAACC$_T^A$ATGTAC. Accordingly, two pools consisting of two oligonucleotides each, and complementary to the above sequence, were synthesized. No other contiguous 5 amino acids in the pheromone could be covered with such a limited set of oligonucleotides.

F. Screening of Recombinant Plasmids

A genomic library, made by insertion of partially Sau3A-digested yeast DNA into the BamHI site of YRp7 (32), was screened for presence of α-factor gene clones. E. coli

transformants were grown on nitrocellulose filter paper (Schleicher
and Schuell, BA85) placed on S-agar plates containing 5 g/ml
ampicillin. After 6 hours at 37°C, filters were transferred to
S-agar plates containing 150 g/ml chloramphenicol. After 15 hours

5   of amplification colonies were tested for hybridization using a
modified in situ colony screening procedure (38). ^{32}P-labeled (40)
synthetic oligonucleotides described above were used as hybridization
probes. Filters were hybridized overnight at 42°C in 10mM Tris (pH 7.5),
6mM EDTA, 0.1mM ATP, 1mM sodium pyrophosphate, 0.8M NaCl, 1X Denhardt's

10   solution, 0.5 percent NP-40, and 0.1 mg/ml E. coli tRNA. Filters were
washed 3 times for 20 min. in 6XSSC at 30°. Dried filters were exposed to
Kodak XR-2 X-ray film with Dupont Lightning-Plus intensifying screen at
-80°.

15

G.   Identification of Recombinant Plasmids Containing the α-factor Gene

       Approximately 4500 bacterial colonies containing recombinant
plasmids were tested for in situ hybridization (38) with

20   ^{32}P-end-labeled oligonucleotide pool I (Fig. 1). Twenty-four
plasmids hybridized to varying degrees. Small amounts of plasmid
DNAs were prepared from these 24 colonies by the method of Birnboim
and Doly (20) and tested for hybridization with the same probes
after spotting the DNA samples on a nitrocellulose filter. Two of

25   the 24 plasmids, designated as p51 and p52 respectively, hybridized
strongly and were chosen for further study. The p51 and p52
plasmids also hybridized with the oligonucleotide pool
II.

30   H.   Subcloning of the Hybridizing Sequences

       To characterize the inserts that hybridized with the
synthetic probes, plasmid DNA prepared from the p51 and p52 clones
was subjected to restriction enzyme analysis with EcoRI, SalI,

35   HindIII, BamHI, and PstI. As seen in Fig. 2A, the 2 recombinant

plasmids are quite dissimilar. Only EcoRI an PstI digestions of the two plasmids yielded one common fragment each. In both cases the common fragment is the TRP1 insert and the 1.38 PstI piece is the DNA between PstI sites in the TRP1 and the amp^R genes.

5

The fragments that contained sequences complementary to the probe were identified by the method of Southern (42). Fig. 2B shows that, except in one case, digestion with all 5 restriction enzymes yielded a fragment that specifically hybridized with the probe. No

10 hybridization was seen with any of the fragments produced by restriction of p52 DNA with HindIII.

The smallest restriction fragments that contained sequences complementary to the synthetic probes were the 1.7 kbp EcoRI

15 fragment from p52 and the 1.8 kbp HindIII fragment from p51. These two DNA fragments were isolated from a preparative agarose gel by electroelution and separately ligated to appropriately cleaved plasmid pBR322 (33) DNA. The ligation mixture was used to transform E. coli 294 and the plasmid DNA from the transformants was analyzed

20 by a quick-screen procedure (20). Two transformants, designated p53 and p56, containing the 1.7 kbp EcoRI and 1.8 kbp HindIII fragment inserts, respectively, were analyzed as follows: Plasmid DNA was prepared from p53 an p56 and digested separately with BamHI, ClaI, PvuI, PstI, and SalI. The resulting DNA fragments were separated on

25 a 1 percent agarose gel, transferred to nitrocellulose filter paper (42) and tested for hybridization with ^{32}p-labeled probes. The analysis of the restriction digests and corresponding hybridization patterns of the p53 DNA, the recombinant plasmid containing the 1.7 kbp yeast DNA as an EcoRI fragment, showed that the yeast DNA in

30 this clone contained one SalI and two PstI sites and that the sequence complementary to the probes was included within a 0.5 kbp PstI-SalI fragment. The HindIII fragment of yeast DNA in the clone p56 lacked recognition sites for these enzymes, and the linearized plasmid, resulting from cleavage at single recognition sites for

35 these enzymes in the pBR322 vector, hybridized with the probes.

This plasmid was then digested with a number of additional
restriction endonucleases and the digests were analyzed by the
method of Southern as described above. It was found that the
hybridizing sequences in this plasmid were contained on a 1.3 kbp

5   HindIII-SacI fragment.

The property of growth inhibition of "a" cells by α-factor
was used to test whether or not the pheromone gene contained in the
cloned 1.7 kbp EcoRI and 1.8 kbp HindIII fragments are functional.

10   If an active α-factor pheromone gene were present in a plasmid, it
would be expected significantly more pheromone would be synthesized
in cells containing the multi-copy plasmid than in cells containing
only the chromosomal copy (or copies) of the gene. The enhanced
level of the α-factor could then be detected by an increase in the

15   area of nongrowth in a lawn of responsive "a" cells. The 1.7 kbp
fragment, isolated from EcoRI-digested p53 DNA, and the 1.8 kbp
fragment, isolated from HindIII-digested p56 DNA, were separately
ligated to a pBR322-based vector plasmid which contained the yeast
selectable marker TRP1 and the yeast origin of replication from the

20   2μm yeast plasmid (43). Yeast strain 20B-12 was separately
transformed with these plasmids and with a control plasmid that
lacked DNA sequences coding for the α-factor. The transformants
were then compared for pheromone production. The transformants
containing MFα1 or MFα2 coding sequences on plasmids produced

25   significantly more α-factor than the same strain-transformed with
the control plasmid. We concluded that the 1.7 kbp EcoRI (MFα1) and
1.8 kbp HindIII (MFα2) fragments contain active α-factor pheromone
genes. The result with MFα1 is consistent with that described by
Kurjan and Herskowitz (44), as this gene corresponds to the gene

30   described by them.

I.   DNA Sequence Determination

35   DNA sequence determination was as previously described

0384L

(45). Briefly, DNA sequences were obtained by the chain termination
method (47) using recombinant phages M13 mp8 and mp9 (39) as the
source for single-stranded "template" DNA and a synthetic
oligonucleotide for priming $\underline{E}$. $\underline{coli}$ DNA polymerase I (large
fragment, Boehringer Mannheim) in the presence of α-^{32}P dCTP (400
Ci/mmole, Amersham). Reactions were electrophoresed on 5 percent
polyacrylamide/8M urea "thin" gels (47). Gels were dried onto 3MM
paper (Whatman) and exposed to X-ray film for 2 to 12 hr.

The nucleotide sequences of large parts of the 1.7 kbp $\underline{EcoRI}$
fragment and the 1.3 kbp $\underline{HindIII}$-$\underline{SacI}$ fragment are shown in Fig. 3
and Fig. 4, respectively. The p53 sequence contains an open reading
frame coding for a protein of 165 amino acid residues which carries
4 internal repeat units within its C-terminal half. Each unit
begins with Lys-Arg and ends with the α-factor sequence. Within
each unit the pair of basic residues is separated from the
α-factorby several Glu (or Asp)-Ala dipeptide repeats. The
N-terminal half of the protein starts with a highly hydrophobic
sequence of 22 amino acids which probably represents a signal
sequence for secretion. The 61 amino acid residues between this
hydrophobic sequence and the first repeat unit include 3 possible
recognition sites for N-glycosylation (indicated by bars in figs.
3). The organization of the pheromone gene contained in p53 clone
is identical to the MFα gene recently described by Kurjan and
Herskowitz (44). This gene differs from $\underline{MF}$α1 at.4 positions. It.
contains T (instead of C) residues at positions -8 and -7, and 125
and an A (instead of C) residue at position 604. Because of ;the
difference at position 125 there is a TTA (Leu) rather than TCA
(Ser) codon at amino acid position 42. We have designated the gene
contained in p53 as $\underline{MF}$α1.

A different α-factor gene, $\underline{MF}$α2, is present in the p56
clone. The organization of this gene (Fig 4) is similar, but not
identical, to the $\underline{MF}$α1. The α-factor encoded by this gene is
apparently made as a precursor protein of 120 amino acid residues

containing two copies of the pheromone. One of the α-pheromone tridecapeptides contained in the putative precursor is identical to the pheromone copies encoded by the MFα1 gene, whereas the second copy contains a Gln-->Asn and a Lys-->Arg.

The organization of these precursors is strikingly similar to that of certain mammalian precursors for neuroendocrine peptides. Thus, like the propiomelanocortin (48, 49), proenkephalin (50-52), and prodynorphin (53), the yeast precursors contain multiple peptide units destined for secretion. In all these precursors the secreted unit is contained on the C-terminal half of the precursor. The N-terminal half of the molecules carry possible glycosylation sites. As is the case for the mammalian multifunctional precursors, glycosylation may be involved in the correct processing of the α-factor precursor. However, the actual processing steps for the yeast precursor seem to be unpredictably different from those of mammalian precursor proteins. Whereas the pairs of basic residues (Lys-Arg) providing sites for release directly flank the secreted peptide in the mammalian precursors, cleavage at these sites in the α-factor precursor would release the pheromone units with several additional amino acids at the N-terminus (see figs. 3 and 4). These N-terminus extensions would consist of repeating -X-Ala- sequences in the precursors encoded by both MFα1 and MFα2 genes. Recent experiments (54, 55) indicate that the last step in the processing of the α-factor precursors is the removal of these sequences by dipeptidyl amino peptidases. The bee venom melittin (56) and the frog skin caerulein (57) precursors are apparently processed by similar mechanisms.

J. Construction of a Plasmid for Expression and Secretion of Human Interferon

Although, as discussed above, our DNA sequence data suggest that the α-factor is synthesized as precursor proteins of 165 and 120 amino acids, no such proteins have been described. The

0384L

processing and secretion mechanism of α-factor is not known. Recent
studies, however, with altered α-factor indicate that the last step
in the production of mature α-factor is apparently the removal of
the glu-ala or asp-ala units before the release of the α-factor
oligopeptides of 12-13 amino acids having the basic sequence
H_2N-(Trp)-His-Trp-Leu-Gln-Leu-Pro-Gly-Gln-Pro-Met (or
MetSO)-Tyr-COOH.

The preparation of a plasmid to demonstrate the usefulness
of the α-factor promoter and the α-factor presequences for
expression and secretion of heterologous gene products is outlined
in Fig. 5. The DNA sequences coding for the α-factor presequences
for expression and secretion of heterologous gene products is
outlined in Fig. 5. The DNA sequences coding for the α-factor
peptides were removed from one of the α-factor clones (p53) such
that the resulting plasmid, p57, contained the promoter sequences
and the sequence corresponding to 89 amino acids of the α-factor
"prepro" protein. This sequence was then joined with human
interferon D (IFN-α_1) gene to form plasmid p58. The human
interferon D gene (58) was modified such that DNA sequences
corresponding to Leu-Glu-Phe had been added before the initiating
methionine codon. After modified interferon D gene had been joined
with the α-factor "prepro" and the promoter sequences, these
sequences were isolated and inserted into a yeast-E. coli shuttle
plasmid YEp9T (Fig. 6). The plasmid YEp9T had been previously made
by replacing the EcoRI-SalI fragment in plasmid YEp1PT (59) with the
EcoRI-SalI fragment from pBR322. This plasmid contains the pBR322
(33) DNA needed for its selection and replication in E. coli. In
addition, it contains the yeast TRP1 gene on an EcoRI to PstI
fragment from chromosome IV (34-36) and a yeast origin of
replication on a PstI to EcoRI fragment from the endogenous 2μ
plasmid DNA. These two DNA fragments from yeast allow for its
selection in yeast and for its autonomous replication and
maintenance as a plasmid. The resulting plasmid, p60, with the
indicated orientation of the insert was selected because the 2μ

0384L

origin contains a transcription termination/polyadenylation signal
(37). The DNA sequence at the junction of the α-factor "prepro"
sequence and the modified LeIFN-D gene present in p60 is shown in
Fig. 7. The p60 plasmid was introduced into the yeast strain 20B-12

5   and the trp$^+$ transformants were grown and assayed for interferon
production.

K. Interferon Assay of Growth Medium and Cell Extracts

10

      Individual colonies of the transformants were grown at 30°C
in 20 ml YNB+CAA to an A_{660} of approximately 10. For assay 10 ml
aliquot was centrifuged at 7K rpm for 10 minutes in a Sorval SM24
rotor. Various dilutions of supernate (media) were assayed. The

15   cells were resuspended in 0.5 ml 7M guanidine-HCl containing an
equal volume of glass beads and vortexed for 2 minutes at high
speed. Both the cell lysate and the medium were then diluted into
PBS/BSA (150 mM NaCl, 20 mM sodium phosphate (pH = 7.9), and 0.5
percent bovine serum albumin) for bioassay. Extracts of yeast were

20   assayed for interferon by comparison with interferon standards by
the cytopathic effect (CPE) inhibition assay (26). Up to one
hundred million units of interferon per liter of growth medium was
found. The cell extracts also yielded interferon at the rate of
100 x 10^6 units per liter of culture.

25

L. Purification of Interferon from the Medium

      A single colony of yeast strain p60/20B-12 was grown at 30°C
in 500 ml YNB+CAA to an A_{660} of 2.4. Five hundred ml of this

30   culture was diluted to 5L with YNB+CAA to give an A_{660} of 0.21;
the resultant 5L culture was grown at 30°C until A_{660} = 70. At
this time the 5L culture was harvested by centrifugation at 7,000
rpm for 10 minutes. Ten ml aliquots were withdrawn periodically
during the fermentation to measure optical density, interferon

35   production and secretion. Before assay, each aliquot was

0384L

centrifuged for 5 minutes in a bench-top refrigerated centrifuge to separate the cells from the medium. The medium and cells were assayed as described above (see Fig. 8). Two different fermentations were done. The peak activity of interferon in the
5   media were 3×10^9 and 2×10^9 units per liter, respectively. The interferon activity in the cell extracts were 1×10^9 and 2×10^9 units per liter of culture.

One and a half liters of frozen medium were concentrated and
10   dialyzed against 25mM Tris, 10mM EDTA, pH 8.0 in a 2.5 liter Amicon stirred cell (Amicon 2000) using a YM-5 ultrafiltration membrane to a final volume of 116 ml. A sample of the retentate was sequenced directly. Another sample of the retentate was acetone precipitated and sequenced.
15

One ml of the concentrated medium was precipitated with 4 ml acetone, spun in a microfuge and washed with acetone. The pellet was resuspended in 0.1 percent TFA and further purified by HPLC on a Synchropak RP-P column. The column was eluted with a linear
20   gradient of 0 to 100 percent acetonitrile in 0.1 percent TFA in 60 minutes. A 12 μg sample of purified IFN-αAD was chromatographed as a control. The peaks of absorbance at 280 nm were sequenced.

25   M.   <u>N-terminal Amino Acid Sequence of Interferon from Growth Medium</u>

Sequence analysis was based on the Edman degradation (27). Liquid samples were introduced into the cap of a modified Beckman 890B spinning cap sequencer. Polybrene^{TM} was used as a carrier in
30   the cap (28). Reagents used were Beckman's sequence grade 0.1 molar Quadrol buffer, phenyl-isothiocynate, and heptafluorabutyric acid. Norleucine was added during each cycle with the Quadrol buffer to serve as an internal standard. The presence of PTH-norleucine in each chromatogram aided in the identification of PTH amino acids by
35   retention time. The amino acid sequence analysis showed only one

species of interferon molecule with the NH_2-terminal sequence
NH_2-Glu-Ala-Glu-Ala-Leu-Glu-Phe-Met. The Met results from the
start codon at the N terminus of the interferon gene, and thus the
protein produced contains 7 extra amino acids, three from the

5   construction, i.e., Leu-Glu-Phe (see figure 7) and 4 from the
presequence of α-factor, i.e., Glu-Ala-Glu-Ala. The polypeptide
containing this 7 amino acid N-terminal extension retains interferon
activity.

10

N. Expression and Secretion of Other Heterologous Gene Products

         In the process of testing the utility of the α-factor
promoter and "prepro"-sequence, restriction endonuclease sites were
15   created at the end of the α-factor "prepro"-sequence (see Fig. 7)
such that the promoter and the "prepro"-sequence could be isolated
as a portable restriction fragment. An appropriate plasmid could
then be constructed to test the efficacy of this expression and
secretion system for any heterologous gene containing suitable
20   "sticky" ends. For this purpose an expression plasmid p65, was
constructed as shown in Fig. 9. This plasmid, like YEp9T, contains
the origins of replication for E. coli and yeast as well as
selective markers for selection in each; of these two organisms. It
also contains a convenient EcoRl site for gene insertion so that any
25   gene that is contained on an EcoRI fragment where the first codon of
the gene is immediately preceded by the EcoRI site could be tested
for the synthesis and secretion of the corresponding protein.

         The plasmid p65 was partially digested with EcoRI the linear
30   molecules isolated, and ligated with EcoRI fragments containing
various genes. After transformation of E. coli, plasmids that
contained the inserts in the appropriate orientation were selected.
For expression the fragment must be inserted at the EcoRI site
following the promoter with the 5' end of the gene connected to that
35   site. This orientation creates the junction between the α-factor

signal sequence and the heterologous gene as previously shown for LeIF-D (see Fig. 7).

Table I lists genes that have been thus tested with the α-factor promoter and signal sequence.

TABLE I Expression and Secretion of Other Heterologous Genes Using the α-Factor Promoter and Signal Sequence

Gene	Growth Vessel	Products per Liter	
		Cellular	Medium
Human Interferon γ	·Shake-flask	10^5 units	trace
Human Serum Albumin	Fermentor	25 mg	3 mg
Bovine Interferon α1	Fermentor	100×10^6 U	200×10^6 U
Bovine Interferon α2	Fermentor	400×10^6 U	60×10^6 U
Tissue Plasminogen Activator	Shake-Flask	$20 \mu g$	$20 \mu g$
Rennin	Shake-Flask	$100 \mu g$	trace
Human Insulin-Like Growth Factor	Fermentor	1-5 mg	3 mg

The expression of the first 4 genes was achieved by the insertion of EcoRI fragments into p65 as described above. The genes were obtained by EcoR1 digestion of plasmids containing them as described in US 438,128 (GB 2116566, EP 88622) (BoIFN); US 297,380 (GB 2105343, EP 73646) (HSA); and US 312,489 (GB 2107718, EP 77670) (γIFN) and elsewhere, e.g. in Interferons edited by Merigan, et al., Academic Press, Inc. (1982), Proceedings of the Symposium on "Chemistry and Biology of Interferons: Relationship to Therapeutics", held March 8-12, 1982, Squaw Valley, California; Lawn, et al. Nucleic Acids Research 9, 6103 (1981); Gray, et al., Nature 295, 503 (1982).

Because of the placement of the restriction sites in the t-PA and rennin genes it was not practical to construct expression plasmids directly as above, but a modified approach was taken. The

construction of the t-PA expression plasmid is illustrated in Fig.
10 using plasmid pt-PAtrpl2 (60) to obtain the t-PA gene by excision
with XbaI and BglIII, pBI (described in US 438,236 (EP 88632), supra)
and YEp13 (43). The rennin expression plasmid was assembled in
5 analogous manner using the rennin gene obtained by XbaI-BclI
excision of pRI (described in US 452,227, EP application 83307841.3).
These two plasmids contain the LEU2 gene for selection in yeast.
Therefore an α leu2 yeast strain was used for transformation with
these plasmids. The prorennin expression plasmid pRI was
10 constructed by incubating EcoRl-Pst I cleaved γ-Imm plasmid with 5'
and 3' segments in the presence of T4 ligase. γ-IMM is a pBR 322
derived plasmid described in U.S. Appln. Serial No. 312,489,
(GB 2107718, EP 77670) and Gray, et al., Supra.

15   The 5' end fragment was a ligation of a synthetic fragment
and a fragment derived from clone PFLA, as described below. The
synthetic fragment has the structure:

```
              [Met] [Ala] [Glu] [Ile] [Thr]
    d A A T T C A T G G C A G A A A T A A C A A G
              G T A C C G T C T T T A T T G T T C C T A G d
20   reading direction--->
```

This sequence contains an EcoRI site at the "upstream end" an
ATG start codon, followed by sequences coding for the first four
amino acids in prorennin, and terminates in a BamHl site.
25

The cloned fragment to which this synthetic fragment is ligated
in order to form the 5' end of the gene comprises a Xma-BamHI
fragment of approximately 440 base pairs derived from the 5' end of
the gene.
30

The appropriate 440 bp fragment was obtained from cDNA derived
from unfractionated mRNA, using as primer, dGATCCGTCGAATTCGG, i.e.,
the "primer probe". The cDNA formed using this primer was size
fractionated as set forth above and fragments having more than 1,000
35

C284L

base pairs inserted into the PstI site of pBR322 for cloning. The
resulting clones were selected using both Tth probe and the
primer/probe as probes. Only colonies hybridizing with both were
selected. From 1,280 transformed colonies, about 300 colonies were
5 obtained which showed hybridization with both probes. These were
examined for presence of the 5' portion of the prorennin sequence as
follows:

The results of a series of double digestions using Ava I-Pvu I,
10 AvaI-BamHI, BglIBamHI, and BglI-EcoRI were analyzed. Advantage was
thus taken of the known Pvu I and Bgl I sites, each 125 base pairs
either side of the pBR322 Pst site utilized for insertion of the
cDNA sequence. These digestions provide suitable fragments for
analysis.

15
The desired clone, PFLA, was selected by analysis of acrylamide
gel electrophoresis performed on the above double digests of mini
preps prepared from the identified clones. Plasmids were then
isolated from PFLA clone, double digested with BamH I and XmaI, and
20 the 440 bp fragment recovered by gel electrophoresis.

The "complete" 5' end was then created by a standard ligation
reaction utilizing the synthetic fragment and the PFLA clone
BamHI-XmaI fragment with T4 ligase followed by cleavage with XmaI
25 and EcoRI. The resulting ligated sequences were purified on
acrylamide gel electrophoresis selecting for the appropriate 455
base pair fragment.

The 3' end fragment was prepared in an manner analogous to that
30 used to prepare the PFLA clone. cDNA containing >1000 bp formed
from unfractionated messenger RNA using oligo-dT as primer, was
cloned as above, and colonies selected with Tth probe. Approximately
50 colonies resulted. The desired clone was selected by analyzing
the results of gel electrophoresis formed on plasmid minipreps which
35 were double digested with BamHI/BglI, PvuI/BamHI, EcoRl/BglI, and

0384L

PvuI/EcoR I again taking advantage of the PvuI and BglII sites flanking the PstI insert site. The plasmids from the desired colony were isolated then cleaved with Xma I and Pst I and electophoresed to isolate the 800 bp sequence of the "3 - 375" fragment. The 5"3-375" fragment extends from the Xma site congruent with that from the PFLA fragment, past the end of the gene to the Pst I insertion site.

The transformants carrying various expression plasmids were 10grown in appropriate media. The cultures were fractionated into supernatants and cells. The supernatants (media) and cell extracts were assayed for the expression and secretion of various gene products. The bovine interferon activity was assayed by comparison with interferon standards by cytopathic effects. The amounts of 15other products in the medium and cell extracts were determined by radioimmunoassays. The values listed in the table are the peak activities. We have determined that both cellular and secreted t-PA molecules possess biological activity.

20
O. Characterization of Secreted Bovine Interferon

The bovine interferon-α_1 secreted into culture medium has been purified and characterized as described below.

25
7 liters of culture medium was adjusted to pH 8.02 with sodium hydroxide. This solution was then loaded onto a 2.5 x 18 cm Nugel ACA column pre-equilibrated with 50mM tris, pH 8.0. After loading, the column was washed with 50mM Tris, 1 percent (w/v) PEG 8000, pH 308.0 until A_{280} was approximately zero. The column was then eluted with 100mM ammonium acetate, 2 percent (w/v) PEG 8000, pH 5.0 followed by 20mM glycine, 2.5 percent (w/v) PEG 8000, pH 2.0. The majority of the interferon activity eluted in a single peak during the pH 5.0 elution. The pooled material from the Nugel column (88 35mL, 150M units) was loaded onto a 2.5 x 5.0 cm SE-53 column

pre-equilibrated with 25mM ammonium acetate, pH 5.0. The interferon
activity eluted in a single peak during the sodium chloride
elution. This pool contained 14mg protein, 106M units in 11 mL.
The purity of this material as judged by SDS PAGE was approximately
80-90 percent.

Ten ml of the SE-53 pool was applied to a 2.5 x 18 cm Sephacryl
S-300 column equilibrated in 25mM sodium phosphate, pH 6.0. The
column was eluted with this buffer and the interferon activity
eluted in a single peak. This final pool contains 8 mg protein, 113
M units in 17 mL.

N-terminal sequence analysis of the protein present in the SE-53
pool indicates that the bovine interferon secreted by yeast has been
processed at three distinct sites. The three sites of processing
and the relative amounts of each are as follows:

$$-8 \qquad -6 \qquad -3 \qquad +1$$
...lys-arg-glu-ala-glu-ala-leu-glu-phe-met-cys-his-leu-pro-his...
63 pct. 13 pct.   24 pct.

As shown above, met immediately precedes the N-terminus of the
bovine interferon; the short peptide extensions do not result in
loss of interferon activity.

Production and Secretion of Mature Heterologous Proteins

In both cases (human IFN-α_1, and bovine IFN-α_1) where the
N-terminal amino acid sequence of the secreted polypeptides was
determined, the proteins contained extensions of 2 to 7 amino acids
in addition to the initiating methionine. Although these
polypeptides have biological activity, it would be preferable to
produce and secrete into the growth medium proteins that are
identical to the proteins from the natural sources.

025

In order to produce the interferon molecule that contains
cysteine as the N-terminal amino acid (as is the case with the
natural IFN-α_1) we needed to modify the junction between the
factor "pre-pro" sequence and the IFN-α_1 gene such that the

5   removal of the modified "pre-pro" sequence will result in release of
a mature interferon molecule containing the natural N-terminus. An
outline of the procedure to obtain such a junction is shown in
Figure. 11. A DNA fragment containing the MFα1 promoter and .
"pre-pro" sequence and the modified IFN-α_1 gene was isolated and

10   cloned into the EcoRI site of M13 mp8 (61). Single-stranded DNA
template was prepared from the recombinant phage containing the
insert in the appropriate orientation. This template was annealed
with a phosphorylated oligonucleotide. The synthetic
oligonucleotide is 24 bases long and is complementary to 12 bases
coding for leu-asp-lys-arg near the C-terminus of the "pre-pro"

15   sequence and to 12 bases coding for cys-asp-leu-pro, the first 4
amino acids of natural IFN-α_1. This primer-templated intermediate
was subjected to extension and ligation reaction at 23°C for 2 hours
in the presence of 500 mMdATP, 100 mMdTTP, 100 mMdGTP, 100 mMdCTP,

20   20 mMdATP, 3 units DNA polymerase (Klenow), and 400 units t4 DNA
ligase in 10 mM Tris pH7.4, 50 mM NaCl and 10 mM Mg 504. Then
additional 3 units of DNA polymerase (Klenow) and 400 units of T4
DNA ligase was added and mixture incubated for 2 hours at 23°C
followed by incubation at 14°C for 15 hours. This mixture was used

25   to transform E. coli JM101 (62). The phage plaques were screened .
for hybridization with the[32]P-labeled oligonucleotide. Templated
DNA from 2 positive recombinant phages was prepared and sequenced
using a primer complementary to IFn-α_1 DNA. Double stranded DNA
from one recombinant phage that contained the desired deletion

30   (deletion of 24 nucleotides shown as a loop in Fig. 11) was
prepared. The EcoRI fragment containing the modified junction
between the "pro-pro" sequence and the IFN-α_1 gene was isolated
and ligated to EcoRI cleaved YEP9T. After transformation of E. coli
with the mixture a plasmid (p76) with correct orientation of the

35   insert was chosen for further study.

p76 DNA was prepared from E. coli and used to transform 20B-12
strain (ATCC 20626) of yeast. Ten liter culture of one transformant
was grown and the culture medium was centrifuged to separate the
medium from the cells.

5

500 ml of yeast medium was dialyzed into 25 nM Tris, pH8.0, 10
mM EDTA. The dialyzed media was then run through an immuno affinity
column containing monocloned antibody to nature (LeIFA). After
washing with 24 mM TRIS pH8.0, 10 mM EDTA, the intrferon activity
10 was eluted with 0.2M acetic acid. The majority of interferon
activity was found in Fraction No. 45. 200 µl of this fraction was
subjected to N-terminal amino acid sequence analysis as described
before. The major sequence found was that of natural interferon D.
The first 8 N-terminal amino acids of the protein were:
15 cys-asp-leu-pro-glu-thr-his-ser.

Additional Explanatory Notes re Figures

Figure 2

20
Localization of homology between the α-factor probes and the DNA
fragments from p51 (a) and p52 (b) recombinant plasmids. The two
plasmids were digested with different restriction endonucleases and
then electrophoresed on a one percent agarose gel. The DNA
25 fragments were transferred to nitrocellulose paper and hybridized to
^{32}P-labeled probes. Panel A: Ethidium bromide stained gel.
Panel B: Southern blot. Lanes: 1, EcoRI; 2, SalI; 3, HindIII; 4,
BamHI; 5, PstI. The arrows indicate the two DNA fragments that were
subcloned. The size standards were derived from lambda, YRp7, or
30 pBR322 DNA.

Figures 3

35   Nucleotide sequence of MFα$_1$ gene and its nontranslated 5' and

3' flanking regions. The predicted amino acids sequence of the
pheromone precursor is also shown. The numbers above and below the
sequence denote the positions of amino acids and nucleotides,
respectively. The four copies of the α-factor sequences are
included in the boxed areas. The asterisks indicate differences in
the nucleotide sequence in one or more copies of the α-factor coding
regions. Three potential N-glycosylation recognition sites are
indicated by bars. This gene corresponds to the gene reported by
Kurjan, et al. (44).

Figure 4

Nucleotide sequence of MFα2 gene and its nontranslated 5' and 3'
flanking regions. The underlined amino acids indicate differences
between the two pheromone copies encoded by the MFα2 gene. See
Fig. 3 for other details.

Figure 5

Joining of the IFN-α$_1$ gene with the α-factor promoter and the
α-factor presequence. Since, as shown in Fig. 3, the 1.8 kbp EcoRI
fragment contained the promoter, the entire DNA sequence 5' to the
sequences coding for the mature α-factor was joined to the modified
IFN-α$_1$ gene such that the α-factor presequence and the IFN-α$_1$
protein would be synthesized as a single precursor protein using the
α-factor promoter.

Figure 7

The protein and DNA sequence at the junction of α-factor
"prepro" and the modified IFN-α$_1$ genes. The XbaI and EcoRI sites
at the junction are indicated.

0324L

Figure 8

Cellular and medium IFN-α_1 levels during fermentation. At various time intervals 10 ml culture was removed from the fermentor, centrifuged to separate the cells and medium. Cell extracts were prepared as described above. Interferon levels in the medium and the extracts were determined.

Figure 9

Construction of a yeast/E. coli shuttle vector plasmid for the expression of heterologous genes using the α-factor promoter and signal sequences. The 1.12 kbp EcoRI fragment containing the promoter and signal sequence was isolated from p58 and was inserted into the EcoRI site of YEp9T.

Figure 10

Assembly of plasmid p68 for expression and secretion of tissue plasminogen activator.

Figure 11

Scheme for in vitro deletion mutagenesis. The 24 nucleotides that were deleted at the junction of MF21 "pre-pro" sequence and the modified IFN-α1 gene are shown as a loop in the figure.

Figure 12

Comparison of amino acid sequences of the putative α-factor precursors encoded by MFα1 and MFα2 genes. Various gaps were created to align the sequences with maximum homology.

Bibliography

1. Novick, et al., Cell 21, 205 (1980).

2. Duntze, et al., Science 168, 1472 (1970).

3. Woods, et al., J. Gen. Microbiol. 51, 115 (1968).

4. Cabib, et al., J. Bacteriology 124, 1586 (1975).

5. Farkas, Microbiol. Rev 43, 117 (1979).

6. Moor, Arch. Mikrobiol. 57, 135 (1967).

7. Goeddel, et al., Nature 287, 411 (1980).

8. Goeddel, et al., Nature 290, 20 (1981).

9. Yelverton, et al., Nucleic Acids Research 9, 731 (1981).

10. Goeddel, et al. Nucleic Acids Research 8, 4057 (1980).

11. Wetzel, American Scientist 68, 664 (1980).

12. Wetzel, et al., Biochemistry 19, 6096 (1980).

13. Davis, et al., Proc. Natl. Acad. Sci. (USA) 78, 5376 (1981).

14. Hitzeman, et al., Nature 293, 717 (1981).

15. Kleid, et al., Science 214, 1125 (1981).

16. Lawn, et al., Nucleic Acids Res. 9, 6103 (1981).

17. Weck, et al., Nucleic Acids Res. 9, 6153 (1981).

18. Stotlzer, et al., Eur. J. Biochem. 65, 257-262 (1976).

19. Stotlzer, et al., Eur. J. Biochem. 69, 397-400 (1976).

20. Birnboim, H.C., et al., Nucleic Acids Research 7, 1513 (1979).

21. Hinnen, A., et al., Proc. Natl. Acad. Sci. (USA) 75, 1929 (1978).

22. Backman, et al., Proc. Natl. Acad. Sci (USA) 73, 4174- (1976).

23. Mandel, et al., J. Mol. Biol. 53, 159 (1970).

24. Beggs, Nature 275, 104 (1978).

25. Miller, J.H., Experiments in Molecular Genetics, pp. 431, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

26. Stewart, W.E. II The Interferon System (Springer, New York, 1979).

27. P. Edman, G. Begg, "A Protein Sequencer" Eur. J. Biochem. 1, 80 (1967).

28. Tarr, G.E., et al., Anal. Biochem. 84, 622 (1978).

29. Clernell, et al., Biochemistry 9, 4428 (1970).

30. Holland, et al., J. Biol. Chem. 254, 9839 (1979).

31. Holland, et al., J. Biol. Chem. 255, 2596 (1980).

32. Struhl, et al., Proc. Natl. Acad. Sci. USA 76, 1035 (1979).

33. Bolivar, F., et al., Gene 2, 950113 (1977).

34. Stinchcomb, D.T., et al., Nature 282, 39 (1979).

35. Kingsman, et al., J. Gene 7, 141 (1979).

36. Tschumper, G., et al., Gene 10, (1980).

37. Hartley, J.L. et al., J.E. Nature 286, 860 (1980).

38. Grunstein, et al., Proc. Natl. Acad. Sci. USA 72, 3961 (1975).

39. Messing, et al., Nucleic Acids Research 9, 309 (1981).

40. Maxam, A.M., and Gilbert, W., Methods in Enzymol. 65, 490 (1980).

41. Crea, R. and Horn, T. Nucleic Acids Res. 8, 2331 (1980).

42. Southern, J. Mol. Biol. 98, 503 (1975).

43. Broach, et al., Gene 10, 157 (1979).

44. Kurjan, et al., Cell 30, 933 (1982).

45. Smith, Methods Enzymol. 65, 499 (1980).

46. Bennetzen, J.L., et al., J. Biol. Chem. 257, 3026 (1982).

47. Sanger, et al., Proc. Natl. Acad. Sci. USA 74, 5463 (1977).

48. Mains, et al., Proc. Natl. Acad. Sci. USA 74, 3014 (1977).

49. Nakanishi, et al., Nature 278, 423 (1979).

50. Comb, et al., Nature 295, 663 (1982).

51. Gubler, et al., Nature 295, 206 (1982).

52. Noda, et al., Nature 295, 202 (1982).

53. Kokidani, et al., Nature 298, 245 (1982).

54. Barnes, et al., Berkeley Workshop on Recent Advances In Yeast Molecular Biology (1982): Recombinant DNA, Esposito, et al. Eds., pp. 295-305, Lawrence Radiation Laboratory, University of California, Berkeley, CA.

55. Julius, et al., Cell 32, 839 (1982).

56. Kreil, et al., Eur. J. Biochem. 111, 49 (1980).

57. Hoffman, et al., EMBO J. 2, 111 (1983).

58. Weck, et al., Nuc. Acid Res. 9, 5153 (1981).

59. Hitzeman et al., Science 219, 620 (1983).

60. Pennica, et al., Nature 301, 214 (1983).

61. Messig, J. (1981) pp. 143-153 in Third Cleveland Symposium on Macronolecules: Recombinant DNA, Ed. A. Walton, Elsevier, Amsterdam.

62. Messig, J., Recombinant DNA Technical Bulletin 2, 43 (1979).

5

10

15

20

25

30

35

0384L

Claims:

1. A process for obtaining a protein heterologous to yeast as a product of yeast expression, which comprises:

5

transforming a yeast organism with a yeast expression vehicle comprising the DNA sequence of the promoter for yeast alpha factor operably connected to a DNA sequence encoding a protein . heterologous to the yeast organism;

10

culturing the transformed organism; and
recovering the protein from the culture.

2. A process for obtaining a protein heterologous to yeast as a product of yeast expression, which process comprises:

15

transforming a yeast organism with an expression vehicle comprising the DNA sequence encoding substantially the pre-pro peptide of yeast alpha factor operably connected in translation reading frame to a DNA sequence encoding a mature protein heterologous to the yeast organism;

20

culturing the transformed organism; and
recovering the protein from the culture.

25

3. A process for obtaining a protein heterologous to yeast as a product of yeast expression, processing and secretion, which process comprises:

30

transforming a yeast organism with an expression vehicle comprising the DNA sequence of the promoter operably linked to substantially the pre-pro peptide sequence of the yeast alpha factor.

35

0384L

gene which is operably connected in translation reading frame to a. DNA sequence encoding a mature protein heterologous to the yeast organism;

culturing the transformed yeast organism; and
recovering protein from its supporting medium.

4. A process for secreting a protein heterologous to yeast into the supporting medium, which process comprises:

transforming a yeast organism with an expression vehicle comprising the DNA sequence encoding substantially the pre-pro peptide of yeast alpha factor, operably connected in translation reading frame to a DNA sequence encoding a mature protein heterologous to the yeast organism; and

culturing the transformed organism.

5. A process of Claim 4 wherein said DNA sequences are under the control of alpha factor promoter.

6. A yeast expression vehicle comprising the DNA sequence of the promoter of the yeast alpha factor gene operably connected to a DNA sequence encoding a protein heterologous to the yeast organism.

7. An expression vehicle of Claim 6 which also includes the DNA sequence encoding substantially the pre-pro peptide of yeast alpha factor operably linked in translation reading frame upstream to the DNA sequence encoding a mature protein heterologous to the yeast organism.

8.   A yeast expression vehicle comprising the DNA sequence encoding substantially the pre-pro peptide of yeast alpha factor gene operably connected in translation reading frame to a DNA sequence encoding a mature protein heterologous to the yeast organism.

9.   An expression vehicle of Claims 6 to 8 wherein the DNA encoding heterologous protein s.a.g. encodes for a protein selected from the group consisting of human interferon, bovine interferon, tissue plasminogen activator, and rennin.

10.   An expression vehicle of Claims 6 to 8 wherein the DNA encoding heterologous protein encodes for insulin-like growth factor.

11.   A yeast organism transformed with an expression vehicle of any one of Claims 6 to 10.

12.   A yeast organism capable of producing mature heterologous protein in the supportive medium, as a product of expression with a N-terminus pre-sequence derived from yeast alpha factor DNA, processing of said pre-sequence and secretion of the mature protein into the medium.

13.   The organism of Claim 12 wherein the mature heterologous protein is human insulin-like growth factor.

0384L

1/11

Fig.1.

Carboxy terminus of α-factor:	Gly —	Gln —	Pro —	Met —	Tyr COOH
	GGU (90)	CAA (20)	CCA (32)	AUG (20)	UAC (33)
Possible codons and their usage	GGC (3)	CAG (0)	CCU (3)		UAU (0)
	GGA (0)		CCC (1)		
	GGG (0)		CCG (0)		

Consensus
oligonucleotides: 5'-GG^T_CCAACC^A_TATGTAC

Synthesized
oligonucleotide
pools complemen-   I.   5'-GTACATTGGTTG^A_GCC
tary to above: II. 5'-GTACATAGGTTG^A_GCC

Fig.7.

ALPHA-FACTOR "PRE-PRO" SEQUENCE						MODIFIED IFN-ALPHA$_1$ GENE			
...LYS	ARG	GLU	ALA	GLU	ALA	LEU	GLU	PHE	MET...
...AAA	AGA	GAG	GCT	GAA	GCT	CTA	GAA	TTC	ATG...
...TTT	TCT	CTC	CGA	CTT	CGA	GAT	CTT	AAG	TAC...

                                              XbaI   EcoRI

Fig.2.
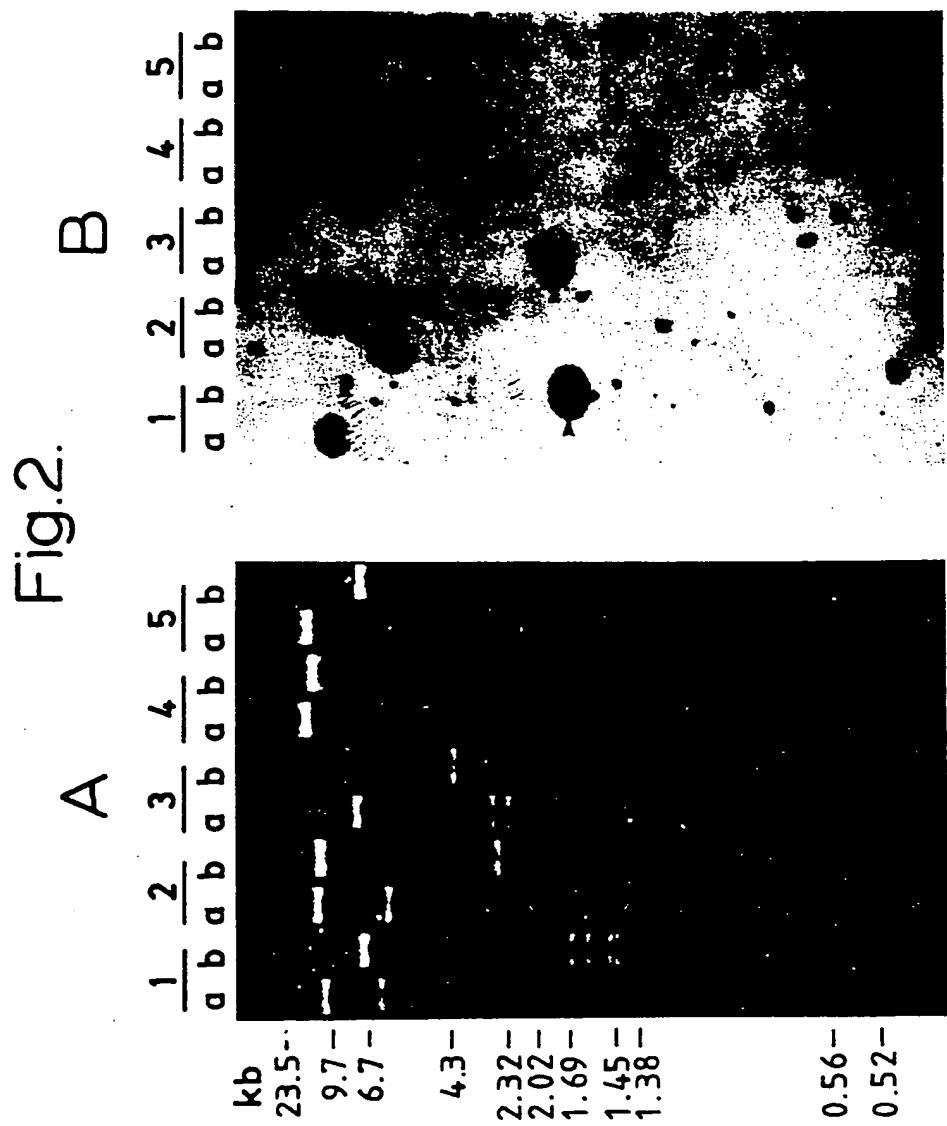
Fig.3.

```
CGACAGTAAATTTTGCCGAATTTAATAGCTTCTACTGAAAAACAGTGGACCATGTGAAAAGATGCATCTCATTTATCAA
        -280              -260            -240              -220

ACACATAATATTCAAGTGAGCCTTACTTCAATTGTATTGAAGTGCAAGAAAACCAAAAAGCAACAACAGGTTTTGGATA
        -200              -180            -160              -140

AGTACATATATAAGAGGGCCTTTTGTTCCCATCAAAAATGTTACTGTTCTTACGATTCATTTACGATTCAAGAATAGTT
        -120              -100            -80              -60
```

```
                                                        1
                                                        Met Arg Phe Pro Ser Ile
CAAACAAGAAGATTACAAACTATCAATTTCATACACAATATAAACGATTAAAAGA ATG AGA TTT CCT TCA ATT
            -40              -20                         1
```

```
                                                20
Phe Thr Ala Val Leu Phe Ala Ala Ser Ser Ala Leu Ala Ala Pro Val Asn Thr Thr Thr
TTT ACT GCA GTT TTA TTC GCA GCA TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA
 20                      40                      60
```

```
                                                40
Glu Asp Glu Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly
GAA GAT GAA ACG GCA CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TTA GAT TTA GAA GGG
 80                      100                     120
```

```
                                                60
Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile
GAT TTC GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG TTT ATA
 140                     160                     180
```

```
                                                80
Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Asp Lys Arg Glu
AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA TCT TTG GAT AAA AGA GAG
 200                     220                     240
```

```
                        ┌─────────────────────────────100──────────────┐
Ala Glu Ala            │Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr│Lys Arg Glu Ala
GCT GAA GCT            │TGG CAT TGG TTG CAA CTA AAA CCT GGC CAA CCA ATG TAC│AAG AGA GAA GCC
                        │        *     * *     *     *                      │
 260                     │       280                    300                 │
```

```
                        ┌─────────────────────────────120──────────────┐
Glu Ala Glu Ala        │Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr│Lys Arg Glu
GAA GCT GAA GCT        │TGG CAT TGG CTG CAA CTA AAG CCT GGC CAA CCA ATG TAC│AAA AGA GAA
                        │        *     * *     *     *                      │
 320                     │       340                    360                 │
```

```
                        ┌─────────────────────────────140──────────────┐
Ala Asp Ala Glu Ala    │Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr│Lys Arg
GCC GAC GCT GAA GCT    │TGG CAT TGG CTG CAA CTA AAG CCT GGC CAA CCA ATG TAC│AAA AGA
                        │        *     * *     *     *                      │
 380                     │       400                    420                 │
```

```
                        ┌─────────────────────────────160────────165────┐
Glu Ala Asp Ala Glu Ala│Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr│End
GAA GCC GAC GCT GAA GCT│TGG CAT TGG TTG CAG TTA AAA CCC GGC CAA CCA ATG TAC│TAA
                        │        *     * *     *     *                      │
 440                     │       460                    480                 │
```

```
GCCCGACTGATAACAACAGTGTAGATGTAACAAAGTCGACTTTGTTCCCACTGTACTTTTAGCTCGTACAAAATACAAT
500              520              540              560

ATACTTTTCATTTCTCCGTAAACAACATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTAC
580              600              620              640

ACATACTTTATATAGCTATTCACTTCTATACACTAAAAAAACTAAGACAATTTTAATTTTGCTGCCTGCCATATTTCAAT
660              680              700              720

TTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTC
740              760              780              800
```

Fig.4.

```
TTCTTCATTGGTACATCAATGCCAGCAACGATGTGCGCATCTGGGCGACGCCTGTAGTGATTGTTTTCAAGGTATCGAG
        -300              -280              -260              -240

CCAAACTATTCATCGTTACTGTTTCAAATATTCAGTTGTTTCAGTACAGAGTCGCCGTGGACCTAGTGAAACTTGGTGT
        -220              -200              -180              -160

CTTTACAGCGCAGAGACGAGGGCTTATATGTATAAAAGCTGTCCTTGATTCTGGTGTAGTTTGAGGTGTCCTTCCTATA
        -140              -120              -100               -80

TCTGTTTTTATATTCTATATAATGGATAATTACTACCATCACCTGCATCAAATTCCAGTAAATTCACATATTGGAGAAA
          -60               -40               -20
```

```
  1                                    10                                    20
Met Lys Phe Ile Ser Thr Phe Leu Thr Phe Ile Leu Ala Ala Val Ser Val Thr Ala Ser
ATG AAA TTC ATT TCT ACC TTT CTC ACT TTT ATT TTA GCG GCC GTT TCT GTC ACT GCT AGT
1                       20                      40                      60
```

```
                                       30                                    40
Ser Asp Glu Asp Ile Ala Gln Val Pro Ala Glu Ala Ile Ile Gly Tyr Leu Asp Phe Gly
TCC GAT GAA GAT ATC GCT CAG GTG CCA GCC GAG GCC ATT ATT GGA TAC TTG GAT TTC GGA
                        80                     100                     120
```

```
                                       50    ▬▬▬▬▬▬▬                        60
Gly Asp His Asp Ile Ala Phe Leu Pro Phe Ser Asn Ala Thr Ala Ser Gly Leu Leu Phe
GGT GAT CAT GAC ATA GCT TTT TTA CCA TTC AGT AAC GCT ACC GCC AGT GGG CTA TTG TTT
                       140                     160                     180
```

```
▬▬▬▬▬▬▬                                70    ▬▬▬▬▬▬▬                        80
Ile Asn Thr Thr Ile Ala Glu Ala Ala Glu Lys Glu Gln Asn Thr Thr Leu Ala Lys Arg
ATC AAC ACC ACT ATT GCT GAG GCG GCT GAA AAA GAG CAA AAC ACC ACT TTG GCG AAA AGA
                       200                     220                     240
```

```
                                       90                                    100
Glu Ala Val Ala Asp Ala│Trp His Trp Leu Asn Leu Arg Pro Gly Gln Pro Met Tyr│Lys
GAG GCT GTT GCC GAC GCT│TGG CAC TGG TTA AAT TTG AGA CCA GGC CAA CCA ATG TAC│AAG
                       │             *      **                              │
                        260                     280                     300
```

```
                                      110                                   120
Arg Glu Ala Asn Ala Asp Ala│Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr│
AGA GAG GCC AAC GCT GAT GCT│TGG CAC TGG TTG CAA CTC AAG CCA GGC CAA CCA ATG TAC│
                           │             *     **                              │
                            320                     340                     360
```

```
End
TGA AAAATGACCCTAAACTACTTCTAAACCCTCTCGATTTCTTTCACGTTCATACAACACCTAGTTTTATTTATTTTC
        380              400              420

TTTTCAATCTGAGTAGTTGAGTTTTCGATCACTCACATAGAACTATTTTTTGCCATTTAAATAAAGTATTCTCTCAAAT
440              460              480              500

GATGCGATACTATAATACTCTTTGCCATATATTACATTCATTCATAAATAGGCTATGTTTCTATATCCGTTTCCGATTC
520              540              560              580

TGTCTGCAAGCAAGGTTCCCTATCATTACCGGATTGTTCACTATGGTTGGAGCTC
600              620              640
```
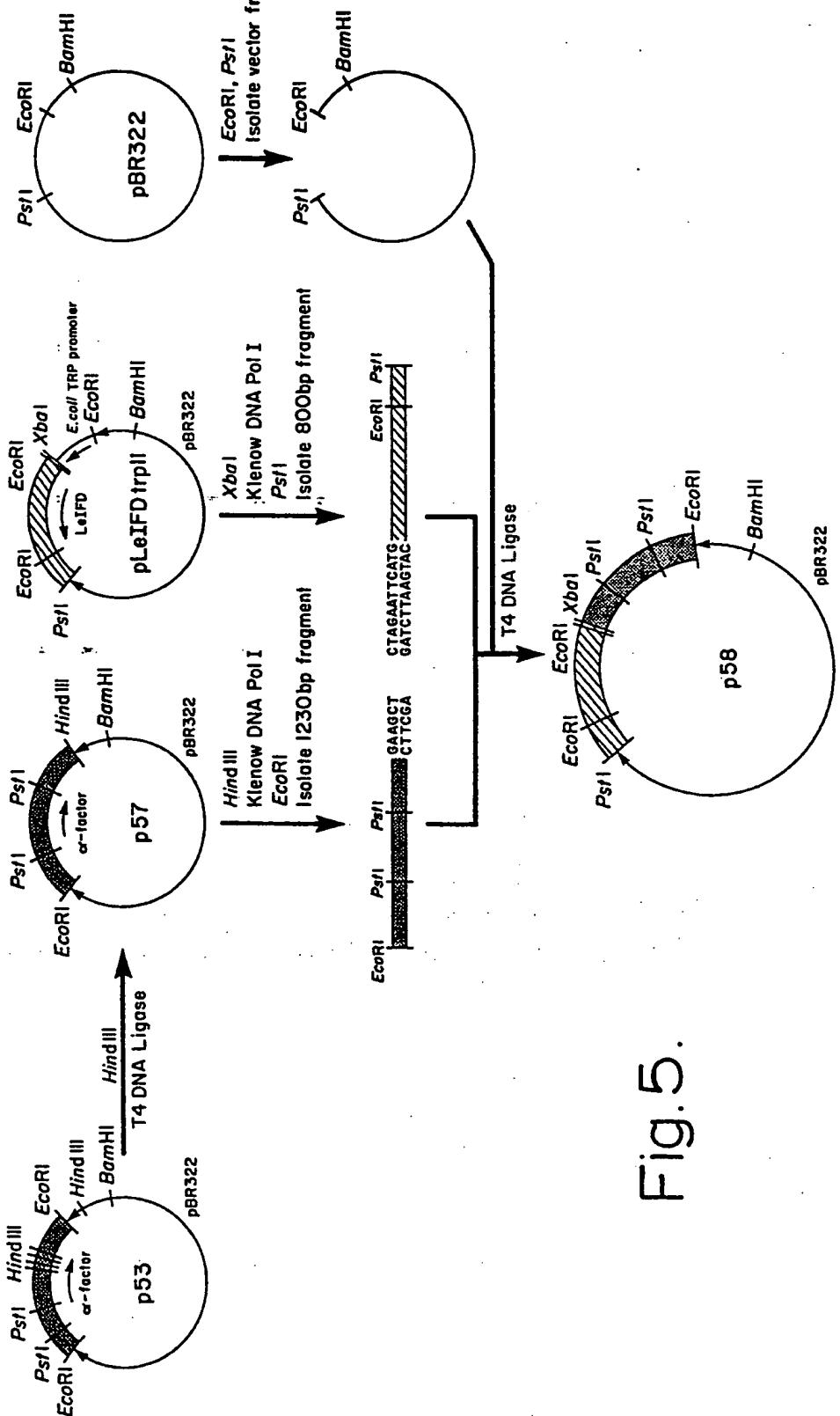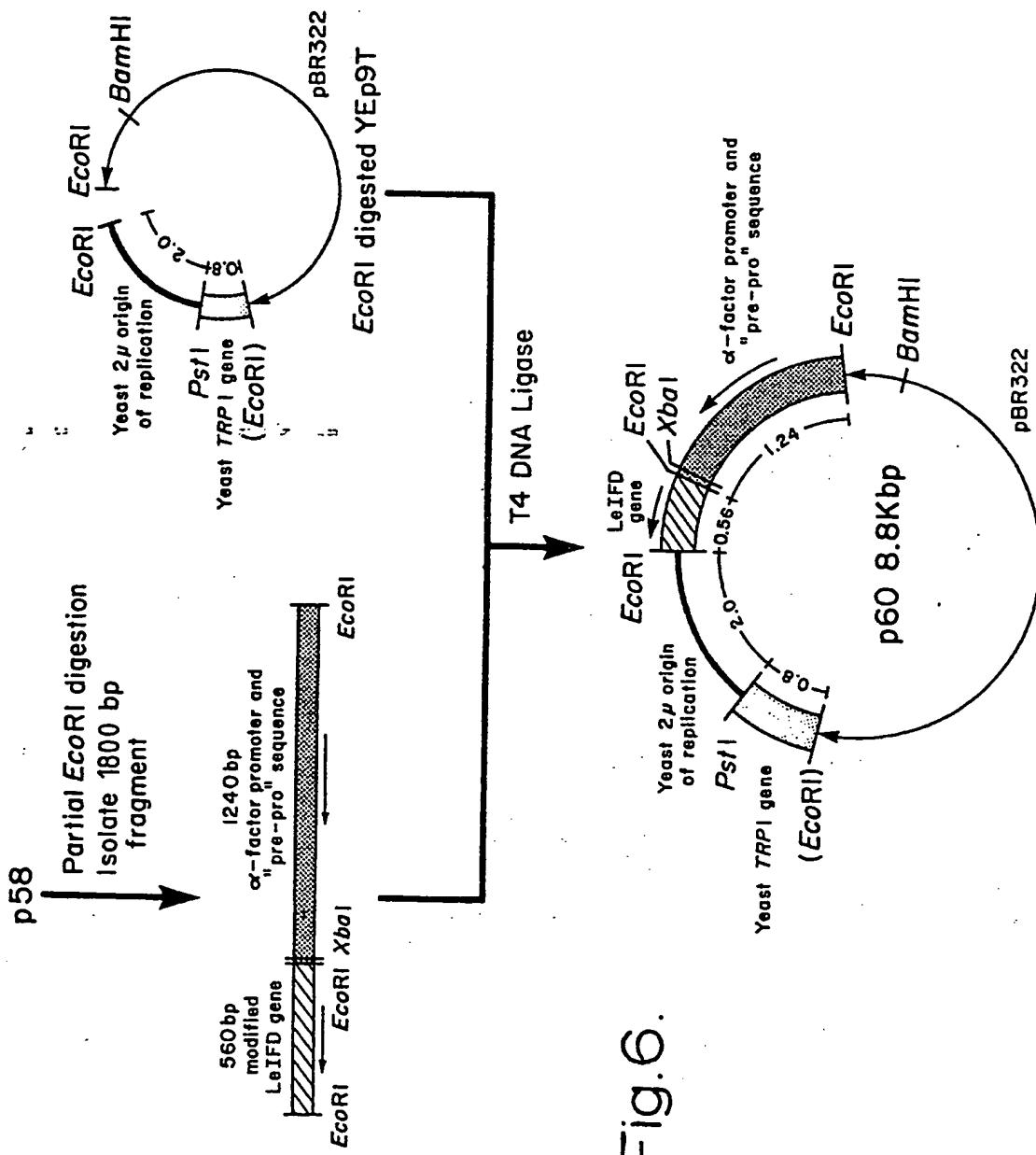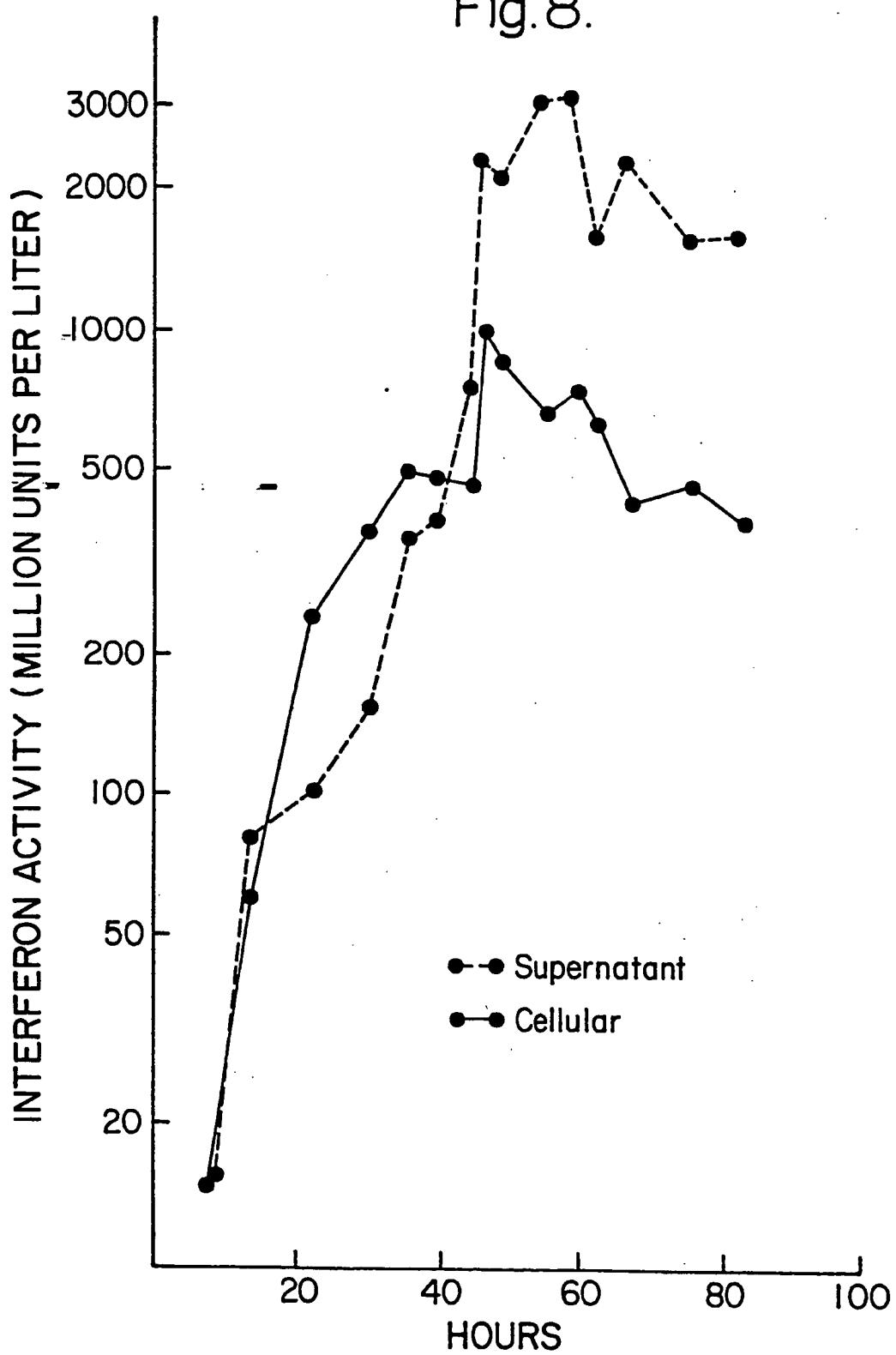
Fig.5.

Fig.6.

0123544

Fig.8.

INTERFERON ACTIVITY (MILLION UNITS PER LITER)

HOURS

●--● Supernatant
●—● Cellular

Fig.9.

p58

EcoRI
Isolate 1240bp
fragment

EcoRI ——————————————— EcoRI

α-factor promoter and
signal sequence

EcoRI

EcoRI

EcoRI digested YEpT9

T4 DNA Ligase

Yeast 2μ origin
of replication

PstI

EcoRI

Yeast TRP I
gene

α-factor promoter and
signal sequence

EcoRI

p65

pBR322
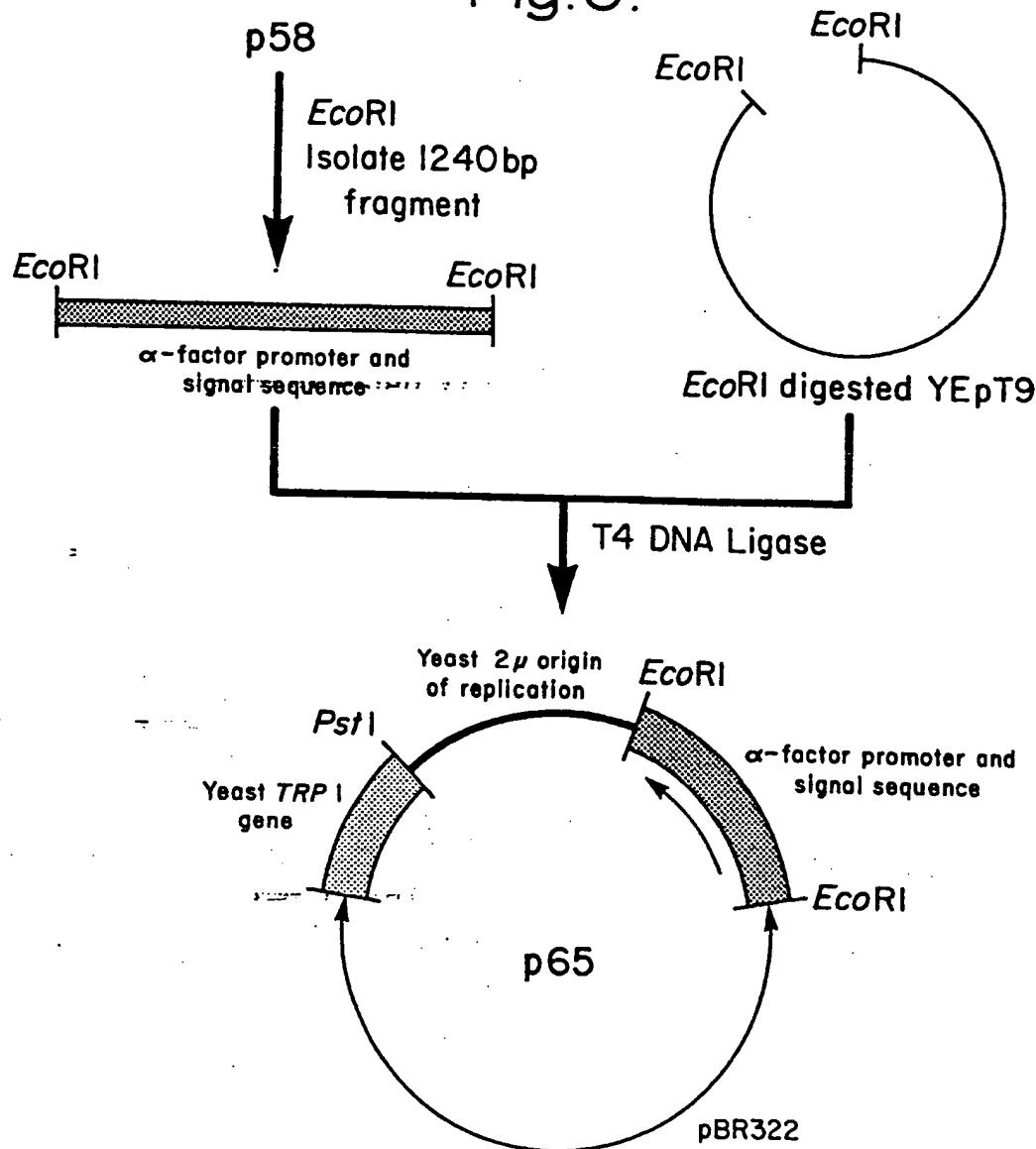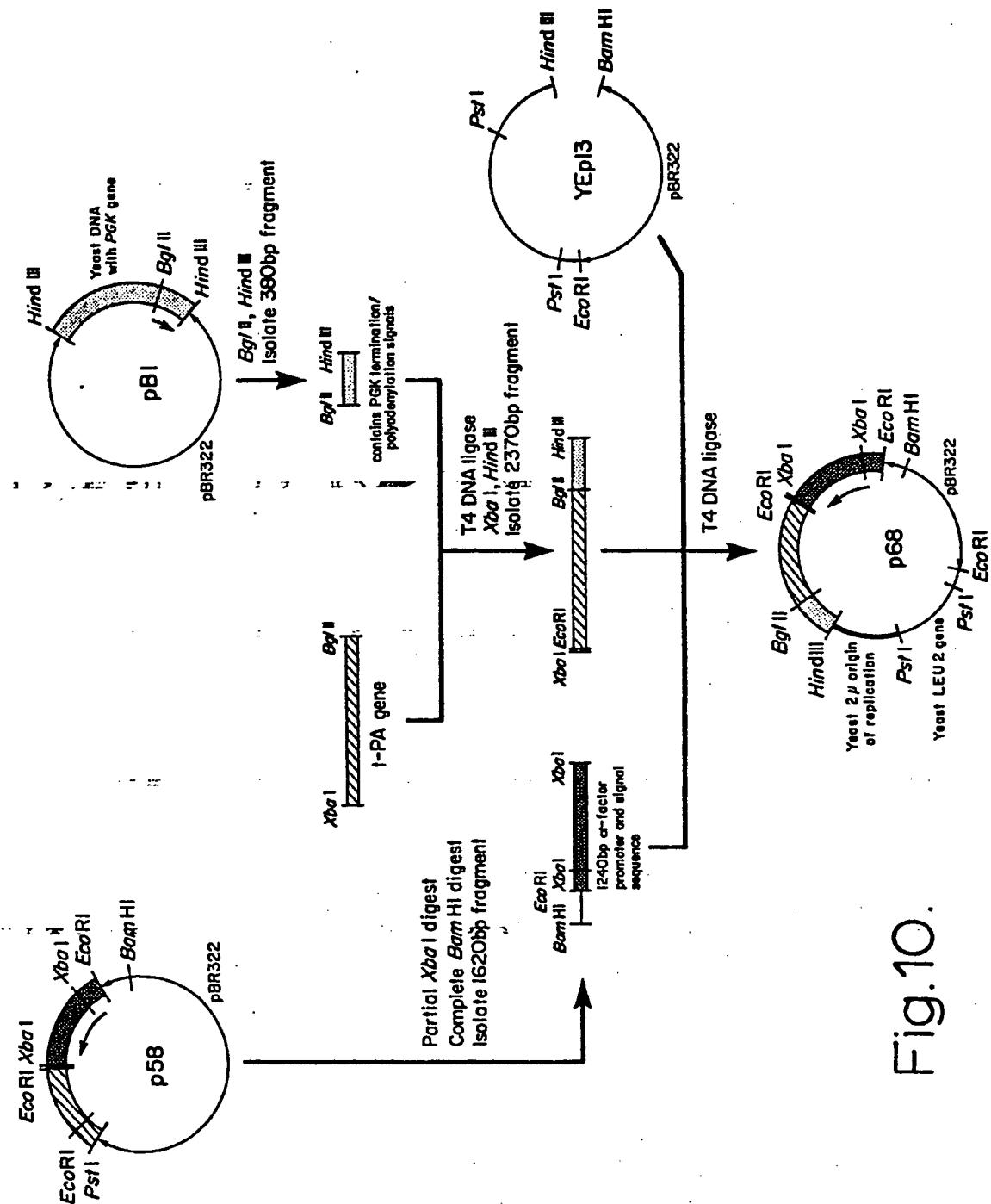
Fig.10.
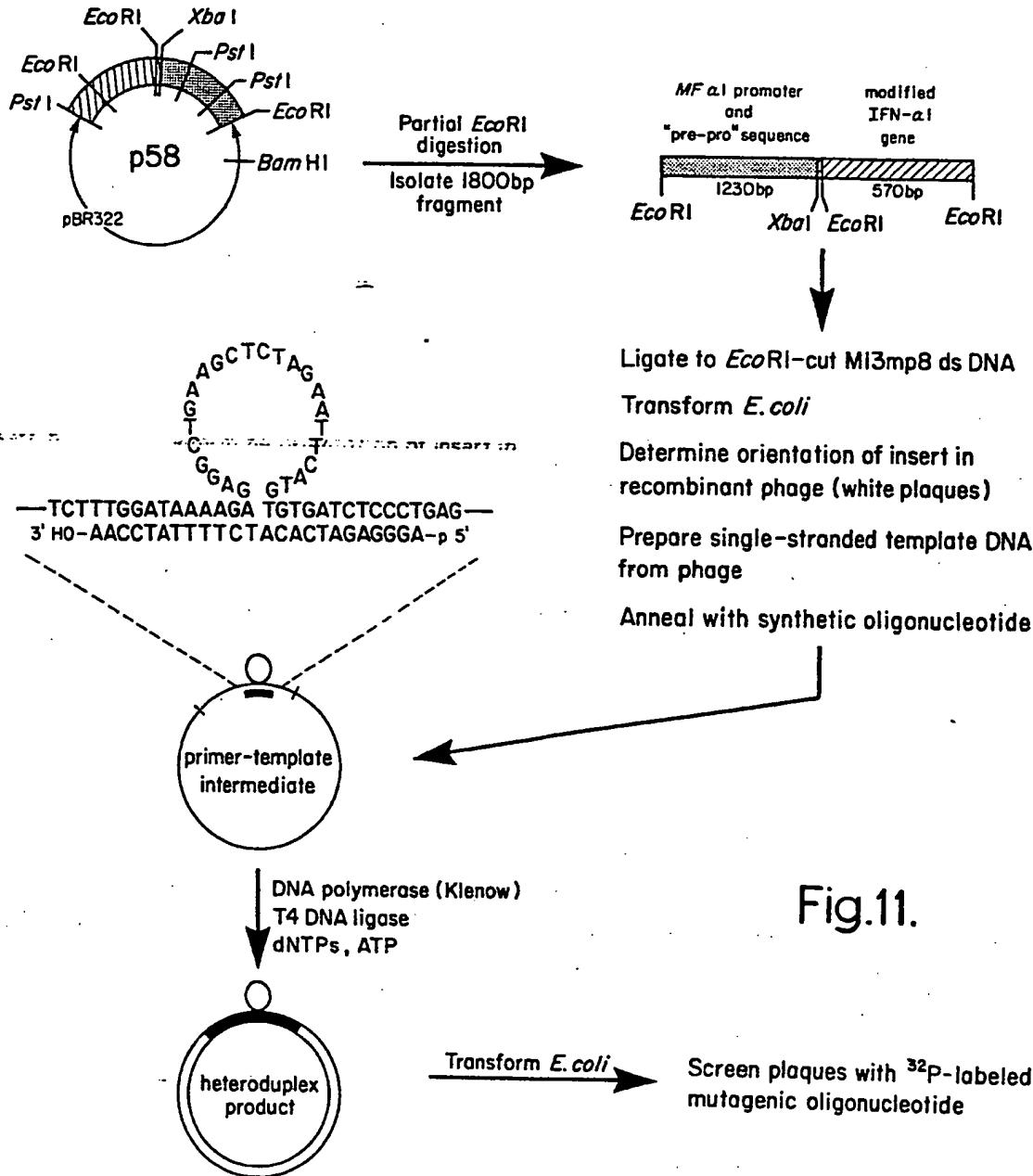
Fig.11.

Fig.12.

	1		10		20
MFα1	MetArgPheProSerIlePheThrAlaValLeuPheAlaAlaSerSerAlaLeuAlaAla				
MFα2	MetLysPheIleSerThrPheLeuThrPhe————————————IleLeuAlaAla				
Consensus	Met——Phe——Ser——Phe——————————————————————LeuAlaAla				

		30			40
MFα1	ProValAsnThr————ThrThrGluAspGluThrAlaGlnIleProAlaAlaValIle				
MFα2	ValSerValThrAlaSerSerAspGluAspAlaGlnIleAlaGlnValProAlaAlaIleIle				
Consensus	————Thr————————————————AlaGln——ProAlaGluAla——Ile				

		50			60
MFα1	GlyTyrLeuAspLeuGluGlyAspPheAspValAlaValLeuProPheSerAsnSerThr				
MFα2	GlyTyrLeuAspPheGlyGlyAspHisAspIleAlaPheLeuProPheSerAsnAlaThr				
Consensus	GlyTyrLeuAsp————GlyAsp————Ala——LeuProPheSerAsn——Thr				

		70			80
MFα1	AsnAsnGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGluGluGly				
MFα2	AlaSerGlyLeuLeuPheIleAsnThrThrIleAlaGluAlaAlaGluLysGluGlnAsn				
Consensus	————GlyLeuLeuPheIleAsnThrThrIleAla————Ala——LysGlu——				

		90			100
MFα1	ValSerLeuAspLysArgGluAlaGlu——————AlaTrpHisTrpLeuGlnLeuLysPro				
MFα2	ThrThrLeuAlaLysArgGluAlaValAlaIleAspAspAlaTrpHisTrpLeuAsnLeuArgPro				
Consensus	————Leu——LysArgGluAla————————AlaTrpHisTrpLeu——Leu——Pro				

		110			120
MFα1	GlyGlnProMetTyrLysArgGluAlaGluAlaGluAlaTrpHisTrpLeuGlnLeuLys				
MFα2	GlyGlnProMetTyrLysArgGluAlaAsnAlaAspAlaTrpHisTrpLeuGlnLeuLys				
Consensus	GlyGlnProMetTyrLysArgGluAla——AlaTrpHisTrpLeuGlnLeuLys				

		130			140
MFα1	ProGlyGlnProMetTyrLysArgGluAlaAspAlaGluAlaTrpHisTrpLeuGlnLeu				
MFα2	ProGlyGlnProMetTyr				
Consensus	ProGlyGlnProMetTyr———————————————————————————				

		150			160
MFα1	LysProGlyGlnProMetTyrLysArgGluAlaAspAlaGluAlaTrpHisTrpLeuGln				
Consensus	———				

	168				
MFα1	LeuLysProGlyGlnProMetTyr				
Consensus	————————————————				